Molecular Recognition between Oligopeptides and Nucleic Acids: DNA Binding Specificity of a Series of Bis Netropsin Analogues Deduced from Footprinting Analysis

Koren L. Kissinger,† James C. Dabrowiak,† and J. William Lown*,‡

Department of Chemistry, Syracuse University, Syracuse, New York 13244-1200, and Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

Received August 8, 1989

Ref. # 9

A series of tether-linked bis netropsins have been synthesized in order to assess the phasing problem, which arises because of the lack of dimensional correspondence between oligopeptides and oligonucleotides in DNA binding characteristics. The consequences of incorporating variable-length flexible and rigid tethers [poly(methylene), *Z* and *E* ethylene, *m*- and *p*-phenylene] between the two netropsin-like moieties on the DNA binding properties were assessed by DNase I footprinting. The conformational freedom associated with two netropsins linked by a flexible methylene tether allows ligand binding in both a mono- and bidentate fashion, with bidentate binding requiring a minimum linker length of $(CH_2)_3$. For compounds possessing rigid tethers, for example, cis and trans ethylene moieties, the cis geometry excludes bidentate ligation while the trans structure favors it. Bis netropsins possessing aryl linking groups have reduced DNA binding affinities. This is most plausibly due to the aryl groups, which are not coplanar with the netropsin moieties, thus blocking the ligand from penetrating deeply into the minor groove of DNA.

Introduction

A number of strategies have been used to target specific sites on DNA (*1*). For single-stranded DNA, oligonucleotides complementary to certain sequences on DNA have been synthesized and characterized. Certain of these oligonucleotides have been equipped with DNA-damaging activity by incorporation of chemical cleaving or alkylating agents into their molecular structure (*2*, *3*). In an effort to minimize attack by nucleases present in the cell and, thus, enhance the potential of these new agents as useful drugs, compounds having modified stereochemistry about phosphorus (*4*) as well as an altered configuration about the anomeric carbon atom of the sugar moiety (*5*, *6*) have been made. In addition to single-stranded DNA, efforts have been made to identify ligands that attack or bind to double-stranded DNA. Among these are single-stranded oligonucleotides, which, under certain conditions, will form a triple helix with duplex DNA (*7*). Although triple-helix formation involves the wider major groove of DNA, the bulk of the small ligands studied to date attack DNA via its minor groove (*1*).

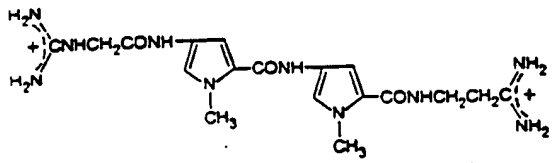
Netropsin and distamycin are *N*-methylpyrrole oligopeptides (Figure 1) which bind in the minor groove of DNA and have high specificity for AT sites. Single-crystal X-ray analyses of the DNA complexes of the drugs bound to small segments of DNA have revealed that hydrogen bonding and van der Waals contacts are the key features in the drug–DNA recognition process (*9–11*). Since the drugs are cationic and the minor groove of DNA in AT-rich regions bears a high negative potential (*12*), the electrostatic potential surface of DNA also plays a role in specificity. Recognition of the importance of the electrostatic effect led to the development of monocationic lexitropsins, or information-reading oligopeptides, in which individual DNA base recognition elements are better able to express themselves since the electrostatic influences are no lc dominant (*13–15*). For example, in an attempt to m the AT specificity of a netropsin-like molecule, we thesized and studied the DNA binding properties series of *N*-methylimidazole oligopeptide agents (*13*, By use of DNase I footprinting methods (*16*), certa these lexitropsins were found to bind in the minor gr of DNA at unique GC-rich sites. This observation, as as NMR evidence (*17*, *18*) led to a binding mecha involving hydrogen bonding between the *N*-me imidazole moiety and the 2-amino group of guanine lo in the minor groove of DNA.

In exploring the structural elements that control ognition, we also studied the chiral oligopeptide antit antibiotics anthelvencin A and dihydrokikumycin B *20*). While the larger tripeptide anthelvencin exhil little enantiomeric selectivity for DNA, the natural isomer of the dipeptidic dihydrokikumycin bound t AT interaction sites with an affinity ~3 times greater its unnatural enantiomer (*20*).

Additional structural elements have been identifi being important in the molecular recognition of c peptides for DNA include the steric hindrance affo by the carboxyl-terminal methylene group in the c peptide antibiotics and, e.g., the guanine 2-NH_2 gro the minor groove of DNA (*21*), which is an impo carrier of biological information (*9*). The fact that steric clash ensures reading of a 3′-AT terminal gro indicated when excision of this methylene permits a 3 recognition and binding as demonstrated by NMR printing (*21*).

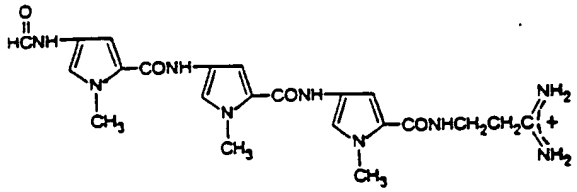
Yet another structural design element emerges fro examination of the DNA sequence selectivity of this bearing lexitropsins. Those structures in which the thi nitrogen is aligned into the minor groove result in G recognition analogous to that of imidazoles. In cont the isomeric structures, in which the larger thiazole s is aligned in to the minor groove, result in GC site dance, presumably for steric hindrance reasons (22

† Syracuse University.
‡ University of Alberta.



1

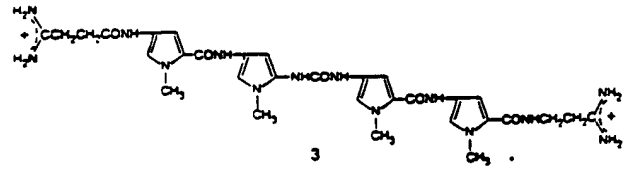
Netropsin



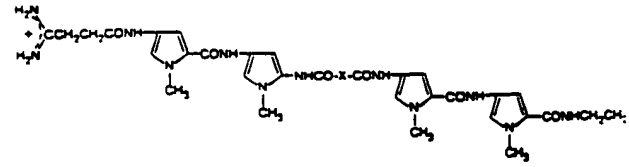
2

Distamycin

Figure 1. The structures of netropsin (1) and distamycin (2) are shown.



3



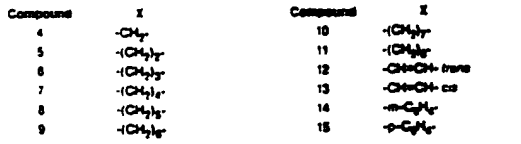


Figure 2. The structures of the bis netropsin ligands used in study are shown.

The ultimate goal in developing lexitropsins, or other groove-selective ligands, as potential gene-control agents is to produce general agents capable of delivery into the cell and of recognizing any predetermined sequence. It can be shown statistically that this requires a run of 15 or 16 base pairs that defines a unique sequence in the human genome (*24*).

One of the central problems in targeting such relatively long segments of DNA using netropsin type molecules concerns the relationship between the repeat distance of a nucleotide unit of DNA and the hydrogen bond and van der Waals contacts generated by an *N*-methylpyrrole peptide. This problem, sometimes referred to as the phasing problem, was recently addressed by Goodsell and Dickerson (*25*). In model studies involving idealized B-DNA coordinates, these workers demonstrated that as the netropsin-like ligand increases in length, the hydrogen bond contacts and van der Waals contacts between the ligand and DNA become seriously out of phase with the spacing between the nucleotides of DNA. Thus, simply increasing the length of *N*-methylpyrrole oligopeptide to gain additional degrees of specificity ultimately results in a severe mismatch between the peptide and DNA which is manifested in reduced binding affinities. This phenomenon probably explains the reduced binding affinities exhibited by certain poly(*N*-methylpyrrole) peptides studied by Dervan and co-workers (*32*).

One way to address the phasing problem is to connect netropsin-like frameworks together by using a tether. This approach, first used by Khorlin and co-workers (*27*), indicated that both netropsin moieties of a tether-linked bis netropsin could bind to DNA. In this paper, we explore the consequences of incorporating variable-length flexible and rigid tethers between two netropsin-like molecules (Figure 2) on the DNA binding properties of the resulting structure. The DNA binding specificities of the compounds were studied by using DNase I footprinting methodology.

Materials and Methods

Melting points were determined on a Fisher-Johns apparatus and are uncorrected. The IR spectra were recorded on Nicolet 7199 FT spectrophotometer, and only the principal peaks are reported. The NMR spectra were recorded on Bruker WH-200 and WH-400 spectrometers. FAB (fast atom bombardment) mass spectra were determined on Associated Electrical Industries (AEI) MS-9 and MS-50 focusing high-resolution mass spectrometers. Kieselgel 60 (230–400 mesh) of E. Merck was used for flash chromatography, and precoated sheets of silica gel $60F_{254}$ Merck were used for TLC. TLC systems were as follows: (covalent peptidic compounds, chloroform/methanol, 9:1; (i ionic compounds with one ionic pair, methanol with some A (iii) for ionic compounds with two ionic pairs, methanol with formic acid.

Synthesis of Linked Bis Oligopeptides. The synthe compounds 3–11 has been reported (*28*).

***N*-[1-Methyl-4-[1-methyl-4-(4-nitrilobutanamido) role-2-carboxamido]-2-pyrrolyl]-*N'*-[1-methyl-2-methyl-2-[[(3-nitrilopropyl)amino]carbonyl]-4-pyrrc amino]carbonyl]-4-pyrrolyl]maleamide.** 3-[[1-Methyl-: methyl-2-amino-4-pyrrolyl)carbamoyl]-4-pyrrolyl]carban propionitrile (*28*) (158 mg, 0.5 mmol) and maleic anhydrid mg, 0.5 mmol) were heated in acetonitrile (5 mL) at 50 ° 3 min. Another portion of 7 (158 mg) was added, and the sol was evaporated to dryness. The residual solid was dissolv DMF (2 mL), DCC (103 mg, 0.5 mmol) was added, an mixture was set aside overnight at room temperature. Two of water were added and the solution was filtered. Then an e of water precipitated the crude product. It was collecte chromatographed on silica gel with chloroform and 15% r anol, providing yellow fractions. These were combinec evaporated, and the residue was recrystallized from aceto give the pure product as a crystalline solid: 100 mg (56.5%) mp 250–2 °C; ^{1}H NMR (DMSO-d_6) δ 2.74 (t, 4 H), 3.54 (q, 3.83 and 3.88 (2 s, 12 H), 6.36 (s, 2 H), 6.84 and 6.88 (2 d, 7.23 (d, 2 H), 7.33 (d, 2 H), 8.36 (t, 2 H), 9.97 (s, 2 H), 11. 2 H); IR (film) 1377, 1438, 1462, 1536, 1630, 1650, 2245, 3270 FAB-MS (*m/z*) 709 MH$^+$. Anal. Calcd for $C_{34}H_{36}N_{12}O_6$: C H, 5.1; N, 23.7. Found: C, 58.0; H, 5.0; N, 23.9.

***N*-[4-[4-(4-Amino-4-iminobutanamido)-1-methylpyr 2-carboxamido]-1-methyl-2-pyrrolyl]-*N'*-[2-[[[2-[[(3-a 3-iminopropyl)amino]carbonyl]-1-methyl-4-pyrrc amino]carbonyl]-1-methyl-4-pyrrolyl]maleamide Dih chloride (13).** The above compound (170 mg, 0.24 mm treated under Pinner reaction conditions as in example 3 28. The completed reaction mixture was evaporated to dr and the residue dissolved in ethanol. Controlled additi 2-propanol provided selective precipitation of impurities. mother liquor was evaporated, the residue was dissolv methanol, and precipitation with acetonitrile gave pure com (13) as a crystalline solid: 166 mg (85% yield); mp 217 ° NMR (DMSO-d_6) δ 2.75 (t, 4 H), 3.54 (q, 4 H), 3.82 and 3 s, 12 H), 6.41 (s, 2 H), 6.97 (d, 2 H), 7.02 (d, 2 H), 7.23 (d, 7.36 (d, 2 H), 8.30 (t, 2 H), 8.77 (bs, 4 H), 9.08 (bs, 4 H), 1 2 H), 11.34 (s, 2 H); IR (Nujol) 1377, 1463, 1528, 1581, 1637, 3252 cm^{-1}; FAB-MS (*m/z*) 743 (M – HCl – Cl)$^+$. Anal. for $C_{34}H_{44}N_{14}O_6Cl_2$: C, 50.1; H, 5.4; N, 24.0; Cl, 8.7. Four 49.7; H, 5.6; N, 24.3; Cl, 8.9.

***N*-[1-Methyl-4-[1-methyl-4-(4-nitrilobutanamido role-2-carboxamido]-2-pyrrolyl]-*N'*-[1-methyl-2-**

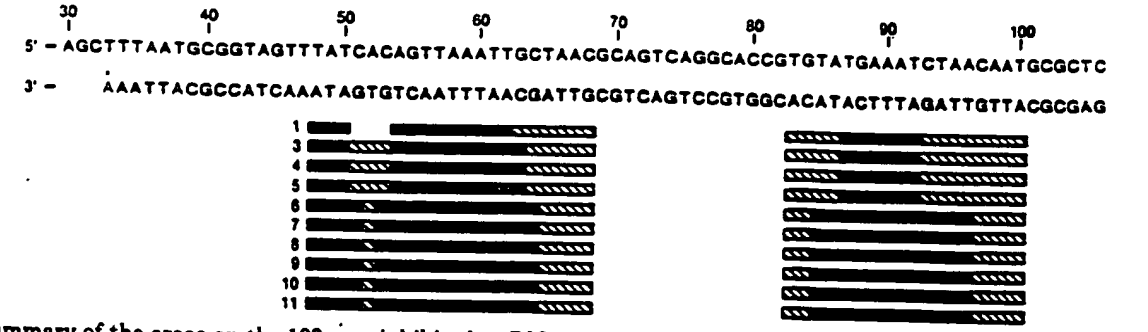


Figure 3. Summary of the areas on the 139-mer inhibited to DNase I digestion by ligands possessing methylene tethers sequence of the 139-mer. The asterisk denotes the position of the radiolabel. Solid bars indicate primary binding sites indicate secondary sites. The uppermost inhibition pattern is that of netropsin (1).

methyl-2-[[(3-nitrilopropyl)amino]carbonyl]-4-pyrrolyl]-amino]carbonyl]-4-pyrrolyl]fumaramide. This compound was prepared by a similar procedure to that described above for the maleic acid derivative in 84.7% yield as a crystalline solid: mp 289–90 °C; ^{1}H NMR (DMSO-d_6) δ 2.74 (t, 4 H), 3.42 (q, 4 H), 3.82 and 3.86 (2 s, 12 H), 6.95 (2 d, 4 H), 7.09 (s, 2 H), 7.22 (d, 2 H), 7.35 (d, 2 H), 8.36 (t, 2 H), 9.98 (s, 2 H), 10.50 (s, 2 H); IR (Nujol) 1328, 1377, 1438, 1464, 1530, 1559, 1586, 1631, 1658, 1692, 2240, 3249 cm^{-1}; FAB-MS (m/z) 709 MH$^+$. Anal. Calcd for $C_{34}H_{38}N_{12}O_6$: C, 57.6; H, 5.1; N, 23.7. Found: C, 57.9; H, 5.3; N, 24.0.

***N*-[4-[4-(4-amino-4-iminobutanamido)-1-methylpyrrole-2-carboxamido]-1-methyl-2-pyrrolyl]-*N'*-[2-[[[2-[[(3-amino-3-iminopropyl)amino]carbonyl]-1-methyl-4-pyrrolyl]-amino]carbonyl]-1-methyl-4-pyrrolyl]fumaramide Dihydrochloride (12).** The title compound was prepared from the above propionitrile intermediate by the Pinner conversion (*28*) described above for the maleic isomer in 58% yield as a crystalline solid: mp 295 °C; ^{1}H NMR (DMSO-d_6) δ 2.65 (t, 4 H), 3.52 (q, 4 H), 3.82 and 3.88 (2 s, 12 H), 6.98 (2 d, 4 H), 7.14 (s, 2 H), 7.24 (d, 2 H), 7.38 (d, 2 H), 8.31 (t, 2 H), 8.95 (bs, 8 H), 10.04 (s, 2 H), 10.64 (s, 2 H); IR (Nujol) 1269, 1377, 1464, 1518, 1579, 1634, 1694, 3230 cm^{-1}; FAB-MS (m/z) 743 (M – HCl – Cl)$^+$. Anal. Calcd for $C_{34}H_{44}N_{14}O_6Cl_2$: C, 50.1; H, 5.4; N, 24.0; Cl, 8.7. Found: C, 49.8; H, 5.6; N, 24.2; Cl, 8.9.

***N*-[1-Methyl-4-[1-methyl-4-(4-nitrilobutanamido)pyrrole-2-carboxamido]-2-pyrrolyl]-*N'*-[1-methyl-2-[[[1-methyl-2-[[(3-nitrilopropyl)amino]carbonyl]-4-pyrrolyl]-amino]carbonyl]-4-pyrrolyl]isophthalamide.** The title compound was prepared by a similar procedure to that described for the maleic and fumaric acid derivatives in 95% yield as a crystalline solid: mp 278–82 °C; ^{1}H NMR (DMSO-d_6) δ 2.76 (t, 4 H), 3.44 (q, 4 H), 3.85 and 3.92 (2 s, 12 H), 6.95 (d, 2 H), 7.14 (d, 2 H), 7.27 (d, 2 H), 7.39 (d, 2 H), 7.68 (t, 1 H), 8.11 (2 d, 2 H), 8.37 (t, 2 H), 8.52 (d, 1 H), 10.02 (s, 2 H), 10.56 (s, 2 H); IR (Nujol) 1260, 1291, 1376, 1400, 1439, 1465, 1514, 1531, 1584, 1638, 2250, 3121, 3307 cm^{-1}; FAB-MS (m/z) 759 MH$^+$. Anal. Calcd for $C_{38}H_{38}N_{12}O_6$: C, 60.1; H, 5.0; N, 22.2. Found: C, 59.8; H, 5.3; N, 22.0.

***N*-[4-[4-(4-Amino-4-iminobutanamido)-1-methylpyrrole-2-carboxamido]-1-methyl-2-pyrrolyl]-*N'*-[2-[[[2-[[(3-amino-3-iminopropyl)amino]carbonyl]-1-methyl-4-pyrrolyl]-amino]carbonyl]-1-methyl-4-pyrrolyl]isophthalamide Dihydrochloride (14).** The title compound was prepared from the above propionitrile intermediate by the Pinner conversion described above for the maleic and fumaric acid structures in 78% yield as a crystalline solid: mp 243–50 °C; ^{1}H NMR (DMSO-d_6) δ 2.65 (t, 4 H), 3.54 (q, 4 H), 3.84 and 3.90 (ws, 12 H), 6.98 (d, 2 H), 7.22 (d, 2 H), 7.25 (d, 2 H), 7.41 (d, 2 H), 7.68 (t, 1 H), 8.14 (2 d, 2 H), 8.29 (t, 2 H), 8.69 (2 s, 5 H), 9.05 (bs, 4 H), 10.06 (s, 2 H), 10.75 (s, 2 H); IR (Nujol) 1261, 1377, 1403, 1463, 1529, 1581, 1643, 1690, 3264 cm^{-1}; FAB-MS (m/z) 743 (M – HCl – Cl)$^+$. Anal. Calcd for $C_{38}H_{46}N_{14}O_6Cl_2$: C, 52.7; H, 5.4, N, 22.6, Cl, 8.2. Found: C, 2.5; H, 5.8; N, 22.9; Cl, 8.5.

***N*-[1-Methyl-4-[1-methyl-4-(4-nitrilobutanamido)pyrrole-2-carboxamido]-2-pyrrolyl]-*N'*-[1-methyl-2-[[[1-methyl-2-[[(3-nitrilopropyl)amino]carbonyl]-4-pyrrolyl]-amino]carbonyl]-4-pyrrolyl]terephthalamide.** The title compound was prepared by a similar procedure to that described for the maleic, fumaric, and isophthalic acid compounds in 99% yield as a crystalline solid: mp 305–10 °C; ^{1}H NMR δ 2.75 (t, 4 H), 3.44 (q, 4 H), 8.84 and 8.90 (2 s, 12 2 H), 7.12 (d, 2 H), 7.26 (d, 2 H), 7.39 (d, 2 H), 8.08 (s (t, 2 H), 10.01 (s, 2 H), 10.50 (s, 2 H); IR (Nujol) 1263, 1438, 1467, 1516, 1556, 1584, 1640, 2250, 3295 cm^{-1}; FAB 759. Anal. Calcd for $C_{38}H_{38}N_{12}O_6$: C, 60.1; H, 5.0, N, 2 C, 60.0; H, 5.3; N, 22.5.

***N*-[4-[4-(4-Amino-4-iminobutanamido)-1-meth 2-carboxamido]-1-methyl-2-pyrrolyl]-*N'*-[2-[[[2-[[3-iminopropyl)amino]carbonyl]-1-methyl-4-p amino]carbonyl]-1-methyl-4-pyrrolyl]terephthal hydrochloride (15).** The title compound was prepare above propionitrile intermediate by the Pinner conv described above for the maleic, fumaric, and isoph derivatives in 64% yield as a crystalline solid: mp 26 NMR (DMSO-d_6) δ 2.65 (t, 4 H), 3.54 (q, 4 H), 3.84 s, 12 H), 6.98 (d, 2 H), 7.16 (d, 2 H), 7.22 (d, 2 H), 7.3 8.10 (s, 4 H), 8.29 (t, 2 H), 8.74 (bs, 4 H), 9.04 (bs, (s, 2 H), 10.59 (s, 2 H); IR (Nujol) 1260, 1377, 1403, 1533, 1553, 1577, 1632, 1671, 1687, 3300 cm^{-1}; FAB-MS (M – HCl – Cl)$^+$. Anal. Calcd for $C_{38}H_{46}N_{14}O_6Cl_2$: 5.4; N, 22.6; Cl, 8.2. Found: C, 53.0; H, 5.6; N, 22.

DNase I Footprinting Studies. The 139-base-pair fragment from pBR322 DNA, the partial sequence o shown in Figure 3, was isolated as previously describ *16*). The fragment was labeled at the 3'-terminal positio [α-^{32}P]ATP in the presence of reverse transcriptase. footprinting reactions were carried out in a total volun containing a final concentration of 37.5 mM Tris (pH mM $MgCl_2$, and 2 mM $CaCl_2$. In the reaction, 2 μL of ca DNA (770 mM in base pairs) was mixed with 2 μL diolabeled fragment. To the 4-μL DNA solution was ac of a binding ligand at a known concentration. After equilibration period, 0.1 unit of DNase I was added and proceeded at 37 °C for 10 min. The final ligand conc in the various footprinting experiments ranged from 1 μM. DNase I digestion was terminated by the addition of a denaturing loading buffer (ULB) containing 0.1 70% urea, and 0.025% each bromophenol blue and xyle One-third of the total volume, 6 μL, was loaded int denaturing polyacrylamide gel and electrophoresed fo imately 2 h at a constant temperature of 55 °C in a therr field gradient, electrophoresis device. Autoradiography w out by using Kodak XAR-5 X-ray film at −20 °C. In certa the autoradiographic data were scanned with a previ scribed microdensitometer (*29*). Ligand concentrati determined optically through the molar extinction coef by direct weighing.

Estimation of Drug–DNA Binding Constants. Dr binding constants were estimated as described previou To 2 mL of Tris–EDTA buffer, pH 8, containing 1.3 μM bromide was added calf thymus DNA to give a final conc of 1.35 μM. The fluorescence was measured after equ for a few minutes, using a Turner Model 430 spectrofluo (Turner-Amsco Instruments, Carpinteria, CA) equipped 150-W xenon lamp, at an excitation wavelength of 525 an emission wavelength of 600 nm. Aliquots of concentra solutions were added, and the fluorescence was m Controls were performed to show that the drugs themse not interfere with the fluorescence measurements at th

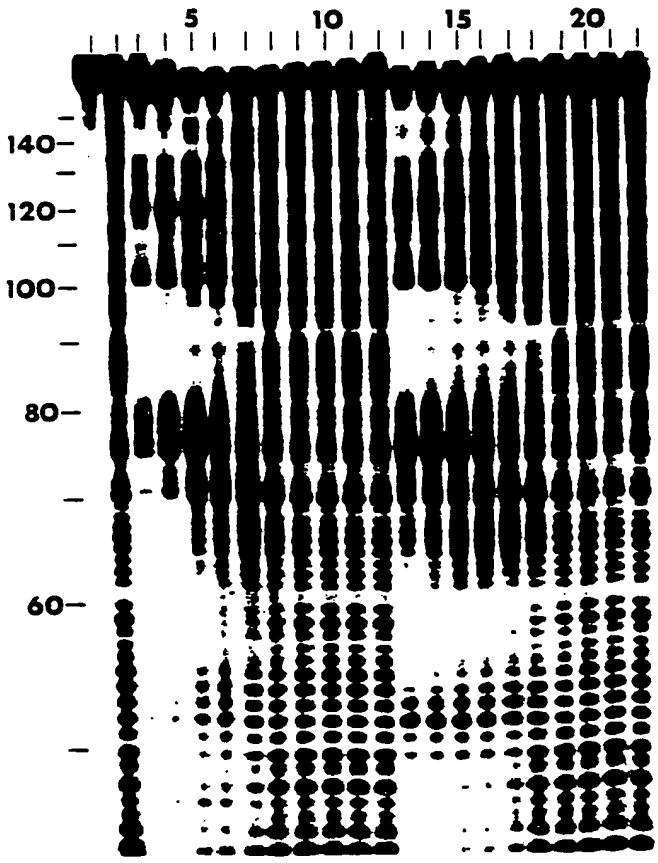


Figure 4. Autoradiographic data taken from footprinting experiments involving methylene-linked ligands 6 [$(CH_2)_3$, lanes 3–11] and 7 [$(CH_2)_4$, lanes 13–21]. Lane 1: 139-mer in the absence of ligand and DNase I. Lanes 2, 12, and 22: DNAse I cleavage of the 139-mer in the absence of ligand. The lane numbers and ligand concentrations in μM are: 3 and 13, 194; 4 and 14, 97; 5 and 15, 48.5; 6 and 16, 24.3; 7 and 17, 11.6; 8 and 18, 5.8; 9 and 19, 2.9; 10 and 20, 1.6; 11 and 21, 0.8.

employed. From a plot of the decreased fluorescence of the ethidium–DNA complex with increasing dose of drug, the concentration of drug needed to reduce the fluorescence by 50% was determined and used to calculate a relative binding constant for the drug, given the binding constant of ethidium to be 10^7 M^{-1} under similar conditions (*31*). This method allows rapid comparison of a series of related compounds, but the individual binding constants may not correspond to values obtained by using other methods (*30*).

Results

Ligands Possessing Methylene Linking Groups. Compounds 3–11 all possess the same basic structure: two netropsin-like portions connected by a methylene chain. The length of the chain ranges from no methylene units to eight methylene units. As the length of the chain increases, the DNase I inhibition regions on the 139-mer enlarge only slightly. Figure 3 shows a summary of the inhibition regions for all nine ligands as well as for that of netropsin deduced from the DNase I footprinting data.

Compounds 3–5, the shortest of the ligands studied, gave patterns identical with that of netropsin. Areas inhibited from enzyme digestion were (5′–3′)ATAAC (50–45), AATTTAAC (62–55), and ATTTCA (92–87). A fourth area of inhibition was located at an AT tetramer, ATAA, around 159–156 (22). The exact size of the region and the nucleotides involved were difficult to assign owing to the low resolution in the higher molecular weight portion of the autoradiogram.

Ligands 6–11 indicated slightly larger protection regions involving nucleotides 51 (G), 54 (T), 86–85 (TA), and 96–63 (TTAG). The footprinting autoradiographic data for 6 and

Table I. Approximate Concentrations at 50% Protec for Ligands with Various Tethers[a]

ligand	tether	concentratio
3	CH_2	3
7	$(CH_2)_4$	3
11	$(CH_2)_8$	3
12	*trans*-CH_2=CH_2	2
13	*cis*-CH_2=CH_2	2
14	*m*-C_6H_4	11
15	*p*-C_6H_4	24

[a] Values were determined by visual inspection of the footp autoradiographic data.

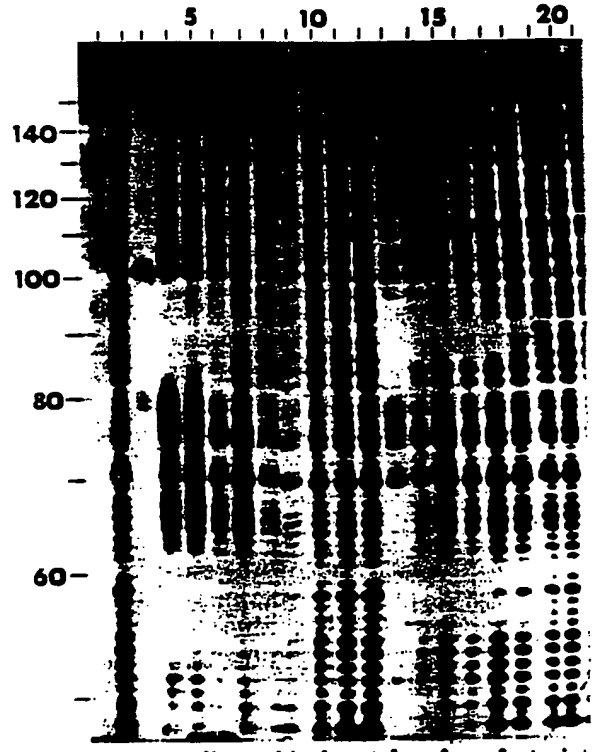


Figure 5. Autoradiographic data taken from footprint periments involving the trans (lanes 3–11) and cis (lanes olefin-linked ligands, 12 and 13. The lane assignments and concentrations are identical with those given in Figure

7 are shown in Figure 4. Comparative DNase printing data for netropsin on this 139-base-pair rest fragment may be found in ref 13 and 14. Second hibition patterns also occurred around the prima involving many AT trimers. Specifically involve nucleotides 54–51 (TGTG) located between two pr netropsin binding sites, 68–64 (GTTAG), 86–83 (T and 101–93 (CATTGTTAG). The 50% protection as visually estimated from the autoradiographic da curred at ~3 μM for all ligands of the series (Ta

Ligands with Olefin Linking Groups. Mark ferences found between the inhibition patterns of th and cis isomers, 12 and 13, respectively (*13*, *14*). pound 12 showed protection patterns encompassin sections of the fragment. Nucleotides 62–48 TAACTGTGAATA) comprised one inhibition are nucleotides 101–81 (CATTGTTAGATTTCATAC comprised the other. No secondary sites could be de for 12. At the highest ligand concentrations studie compound exhibited only two bands on the autorad (Figure 5), corresponding to the single- and d stranded (at ~100) 139-mer.

Conversely, the cis isomer, 13, exhibited a ve quence-specific pattern much like that of netrops *14*). Primary inhibition regions involved nucleotides 50–48 (ATA), 62–56 (AATTTAA), and 92–87 (AT

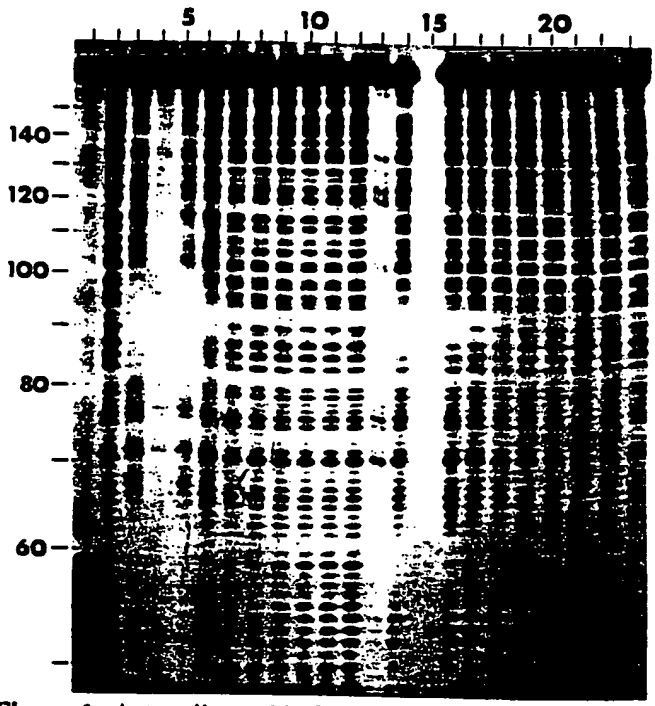


Figure 6. Autoradiographic data taken from footprinting experiments involving the aryl-linked ligands 14 (lanes 3–11) and 15 (lanes 13, 14, and 16–22). Except for the fact that lane 15 was not loaded, the lane assignments and ligand concentrations are identical with those given in Figure 4.

while nucleotides 55–51 (CTGTG), 68–63 (GTTATC), and 101–93 (CATTGTTAG) were included in secondary sites. The midpoint of the titration of the strongest sites occurred at ~2 μM (Table I) for both ligands.

Ligands Containing Aryl Linking Groups. The most prominent feature about the autoradiogram of 14 and 15 (Figure 6) is the low binding affinities of the ligands for DNA. Compound 14, the ligand with the meta-disubstituted aryl ring, showed 50% inhibition at ~11 μM. This value is 4–5 times greater than the amount of ligand necessary for 3–13 to produce similar protection (Table I). Initial regions of protection included nucleotides 50–48 (ATA), 62–55 (ATTTAAC), and 92–87 (ATTTCA). Secondary sites occurred at nucleotides 86–84 (TAC) and 96–93 (TTAT). Compound 15, which has the para-disubstituted aryl ring, showed even lower affinity for DNA than its meta counterpart. The halfway point of the titration with 15 occurred at ~24 μM. Areas of inhibition for 15 were similar to those of 14 (Figure 7).

Discussion

Ligands Possessing Methylene Chains. Lexitropsins 3–11 were synthesized to accomplish two tasks. They were designed to traverse long sequences of DNA but at the same time maintain strong binding. The oligo(*N*-methylpyrrolecarboxamide) molecules synthesized by Youngquist and Dervan (*32*) were successful in binding to longer segments of DNA. Unfortunately, the difference in the hydrogen bond periodicity of the latter ligands and the DNA decreased the binding affinities of the longer peptides. In an effort to overcome the phasing problem, compounds with two pyrrole domains connected by a flexible methylene chain were designed and synthesized. It was theorized that the nonbinding tether would adjust its length to allow both pyrrole portions of the ligand to simultaneously bind to DNA in a bidentate fashion (Figure 7). Since the pyrrole units have high affinities for AT sequences, a potential DNA binding site for the ligands consists of either two sufficiently long AT regions separated by a short nonbinding sequence or a long cor tract of AT base pairs.

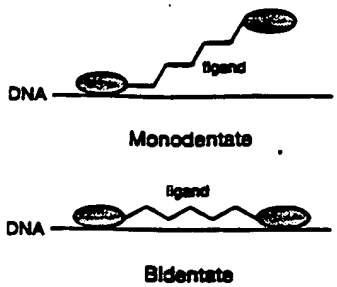


Figure 7. Mono- and bidentate ligation of a bis netropsir is shown.

An area of the 139-mer which contains two binding sites separated by a relatively short non sequence is located at nucleotides 62–46. The se of the site is 5′-AATTTAACTGTGATAAA-3′. compounds bind to this sequence such that the moieties from a single ligand occupy both AT-ri in the sequence, the methylene linker will be pos over the nonbinding pentamer at nucleotides (CTGTG). In this case, enzymatic cleavage will hibited over the entire region.

As is evident in Figures 3–5, the region from nuc 62 to 46 does not titrate simultaneously. Inhibit curred at the AT heptamer (62–56) and pentamer before they occurred at the central pentamer. T sults indicated that only one of the two binding d of the ligand interact with DNA at any one time a monodentate ligation is taking place (Figure 7). Si pentameric sequence is longer than the longest moiety in any of the ligands, i.e., compound 11, havior is expected.

The initial inhibition regions for compounds 3 generally broader than those observed for netropsir monodentate binding allows the unligated portio tend in either direction away from the site of DNA steric effects between the ligand and DNase I shou to either "side" of the site. This effect would block by the enzyme and give rise to inhibition patterns broader than those of netropsin.

Quantitative footprinting studies involving ne and the 139-mer revealed that the primary bindin 89–92 on the restriction fragment has a binding c of ~10^8 M^{-1} (*33*, *34*). Secondary netropsin sites have sequences of the type $(A{\cdot}T)_2(G{\cdot}C)$, have lower constants in the range 10^5–10^6 M^{-1}. Interestingly, lexitropsins 6–11 exhibited initial inhibition patter were larger than those of netropsin and compour in the region 100–85 of the fragment. This region c one strong netropsin site, an AT tetramer, which is by two weaker sites. Thus, it appears that the having the longer linkers may be engaged in bi ligation by spanning a G-C base pair which separa AT tetramer (strong site) from flanking AT trime like netropsin, the binding of a bis lexitropsin in odentate fashion to the high-affinity tetramer re a local high concentration of the unbound portion ligand in the vicinity of an AT trimer. This may ca adjacent AT trimer to bind the remaining portion bis lexitropsin at ligand concentrations not norm sociated with site loading of AT trimers (*33*, *34*). I with shorter linking groups, e.g., 3–5, are not cap spanning the intervening G-C base pair and thus only bind in a monodentate fashion to the tetramer in this region.

It may be argued that the best way to show bi

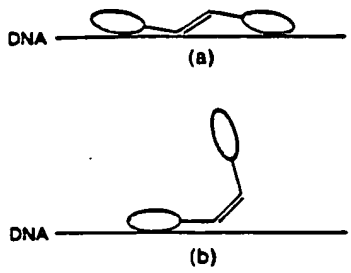


Figure 8. Schematic representations of the cis olefin 12 (a) and trans olefin 13 (b) bound to DNA are shown.

binding would be to employ a restriction fragment having a contiguous sequence of A·T base pairs sufficiently long to bind the entire ligand in a bidentate fashion. However, in this case it becomes difficult to distinguish between monodentate and bidentate binding. This is because the binding constants for all possible "half-sites" in the long sequence are likely to be similar and monoligation on the different restriction fragments present in the mixture will give the impression that a single ligand is binding in a bidentate fashion to the entire sequence. Thus, the aforementioned observation that a weak netropsin site adjacent to a strong site, positions 100–85 of the 139-mer, binds a bis netropsin as well as the strong site suggests that the argument involving a local concentration effect is correct and that bidentate ligation is occurring.

Ligands with Olefin Linkers. Compounds 12 and 13 were designed to investigate the effects of the presence of a rigid linker as well as possible cis/trans effects on bidentate–ligand binding. The autoradiogram from the footprinting experiments involving these compounds (Figure 5) showed dramatic differences in inhibition patterns for the two ligands. Molecular models show that monodentate binding of the trans isomer forces the unbound portion of the lexitropsin to be in the vicinity of DNA (Figure 8). This effect is greater than with 3–11, which have flexible linkers and hence greater degrees of conformational freedom. The additional steric constraints of 12 may be the reason why the initial inhibition region on the 139-mer for the compound appeared to be slightly larger than the analogous methylene-linked systems (Figures 3 and 4). Similar to the argument advanced for bidentate binding of 6–11, one part of lexitropsin, 12, occupies a high-affinity site, while the other portion may be "bound" to a flanking sequence irrespective of the bases present in the sequence.

The patterns resulting from the cis olefin ligand, 13, were not as broad as those for the trans olefin. In fact, they resemble those of netropsin. The relatively narrow patterns for 13 can be explained by considering the geometry about the cis olefin linking group. The optimal angle between the two netropsin moieties of the bis netropsin is not ~180° as in the case of 12, but is probably closer to ~90° (Figure 8). Molecular models show that 13 can bind to DNA in only a monodentate fashion with one netropsin moiety bound to the polymer and the other directed away from the minor groove of DNA. Because the unbound portion of the ligand is not in close proximity to the groove floor, the enzyme is able to cleave DNA near to the nucleotides actually involved in the binding interactions, thus yielding narrow inhibition patterns.

Ligands Containing Aryl Linking Groups. The binding affinities of the aryl-linked compounds, 14 and 15, were much lower than the affinities of the previously discussed ligands. Although 50% protection points for 3–13 occurred at concentrations ranging from 2 to 3 μM, comparable protection was not seen with the aryl-linked compounds until 11–24 μM (Table I). The reduce finities of these compounds are probably the resu nonbonding interactions present in the structures. to steric interference between the amide carbonyl an aryl ortho hydrogens, the aryl rings of 14 and 1 probably perpendicular to the two adjacent amide link This also makes the aryl ring approximately perpendi to the pyrrole rings of both netropsin moieties. Whe ligands bind to DNA, the amide and the C-3 hydr of the pyrrole units can be directed in toward the r groove just as they are in the netropsin–DNA com However, the perpendicular orientation of the aryl blocks deep penetration of the ligand into the minor g of DNA, thereby reducing binding affinities through effects. While there has been a proposal that the o plane piperazine ring of Hoechst 33258 serves as a c minant for a widened minor groove, i.e., a GC site (*35* has been disputed in a reexamination of the Hoechst complex with d(CGCGAATTCGCG)$_2$ (*36*). Cert there is no evidence from the present footprinting dat this is the case for 14 and 15.

In summary, the conformational freedom associated two netropsins linked by a flexible methylene tether a ligand binding in both a mono- and a bidentate fas In the restriction fragment studied, bidentate ligation place by forcing the half-site with the lower binding stant to accept ligand at the same time that the hal with the higher binding constant binds a bis netr This is probably due to a local high concentration c of the "feet" of the ligand, causing an otherwise wea to exhibit binding. For compounds possessing rigid te for example, cis and trans ethylene mono versus bide binding is influenced by the relative dispositions of th feet of the bis netropsin. For the cis structure it possible for both "feet" to simultaneously bind to However, the trans geometry forces both "feet" c ligand to be near DNA. If one is bound to a high-af sequence, the other may extend over or be bound to a affinity site. Bis netropsins possessing aryl linking g have reduced DNA binding affinities. This is most due to the aryl groups, which are not coplanar wit netropsin moieties, thus blocking the ligand from d penetrating the minor groove of DNA.

The relationship between the tethered netropsin the phasing problem (*37*) was not directly addressed i study. Since the restriction fragment possesses a h geneous sequence and not all sites on the fragment ligand with equal affinities, possible mismatches be hydrogen bond donor and acceptor sites on the DN. ligands cannot be detected in the footprinting Studies involving a more uniform binding substrate poly(dAdT)·poly(dAdT) and the bidentate ligands in progress), would appear to directly address this

Efforts directed toward incorporating these mol recognition parameters in the design of longer lexitr to target general predetermined sequences in the h genome will be reported in due course.

Acknowledgment. This research was support grants (to J.W.L.) from the Medical Research Cou Canada and the Natural Sciences and Engineerin search Council of Canada and by Grant NP-681 (to J from the American Cancer Society.

References

(1) Hurley, L. A., and Boyd, F. L. (1987) Topics in Chemist Drug and Design. *Annu. Rep. Med. Chem.* **22**, 259–268.
(2) François, J.-C., Saison-Behmoaras, T., Chassignol, M., T N. T., Sun, J.-S., and Hélène, C. (1988) Periodic Clea Poly(dA) by Oligothymidylates Covalently Linked to th

Phenanthroline–Copper Complex. *Biochemistry* 27, 2272–2276.
(3) Dreyer, G. B., and Dervan, P. B. (1985) Sequence-Specific Cleavage of Single-Stranded DNA: Oligodeoxyribonucleotide–EDTA–Fe(II). *Proc. Natl. Acad. Sci. U.S.A.*, 82, 968–972.
(4) Blake, K. R., Murakami, A., Spitz, S. A., Glave, S. A., Reddy, M., Ts'o, P. O. P., and Miller, P. S. (1985) Hybridization Arrest of Globin Synthesis in Rabbit Reticulocyte Lysates and Cells by Oligodeoxyribonucleoside Methylphosphonates. *Biochemistry* 24, 6139–6145.
(5) Morvan, F., Rayner, B., Imbach, J. L., Chang, D. K., and Lown, J. W. (1986) a-DNA I. Synthetis, Characterization by High Field ^{1}H-NMR, and Base-Pairing Properties of the Unnatural Hexadeoxyribonucleotide a-[d(CpCpTpTPCpC)] with its Complement b-[d(GpGpApApGpG)]. *Nucleic Acids Res.* 14, 5019–5035.
(6) Thuong, N. T., Asseline, U., Roig, V., Takasugi, M., and Hélène, C. (1987) Oligo(a-Deoxynucleotides) Covalently Linked to Intercalating Agents: Differential Binding to Ribo- and Deoxyribopolynucleotides and Stability Towards Nuclease Digestion. *Proc. Natl. Acad. Sci. U.S.A.* 84, 5129–5133.
(7) Moser, H. E., and Dervan, P. B. (1987) Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation. *Science* 238, 645–650.
(8) Praseuth, D., Perrouault, L., Doan, T. L., Chassignol, M., Thuong, N., and Hélène, C. (1988) Sequence-Specific Binding and Photocrosslinking of a and b Oligodeoxynucleotides to the Minor Groove of DNA via Triple-Helix Formation. *Proc. Natl. Acad. Sci. U.S.A.* 85, 1349–1353.
(9) Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. (1985) The Molecular Origin of DNA–Drug Specificity in Netropsin and Distamycin. *Proc. Natl. Acad. Sci. U.S.A.* 82, 1376–1380.
(10) Coll, M., Aymami, J., van der Morel, G. A., van Boom, J. H., Rich, A., and Wang, A. H.-J. (1989) Molecular Structure of the Netropsin–d(CGCGATATCGCG) Complex: DNA Conformation in an Alternating AT Segment. *Biochemistry* 28, 310–320.
(11) Coll, M., Frederick, C. A., Wang, A. H.-J., and Rich, A. (1987) A Bifurcated Hydrogen-Bonded Conformation in the d(A·T) Base Pairs of the DNA Dodecamer d(CGCAAATTTGCG) and its Complex with Distamycin. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8385–8389.
(12) Lavery, R., and Pullman, B. (1982) The Electrostatic Field of DNA: The Role of the Nucleic Acid Conformation. *Nucleic Acids Res.* 10 (14), 4383–4395.
(13) Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., and Dabrowiak, J. C. (1986) Molecular Recognition between Oligopeptides and Nucleic Acids: Novel Imidazole-Containing Oligopeptides Related to Netropsin that Exhibit Altered DNA Sequence Specificity. *Biochemistry* 25, 7408–7416.
(14) Kissinger, K., Krowicki, K., Dabrowiak, J. C., and Lown, J. W. (1987) Molecular Recognition between Oligopeptides and Nucleic Acids: Monocationic Imidazole Lexitropsins that Display Enhanced GC Sequence Dependent DNA Binding. *Biochemistry* 26, 5590–5595.
(15) Lown, J. W. (1988) Lexitropsins: Rational Design of DNA Sequence Reading Agents as Novel Anticancer Agents and Potential Cellular Probes. *Anti-Cancer Drug Des.* 3, 25–40.
(16) Lane, M. J., Dabrowiak, J. C., and Vournakis, J. N. (1983) Sequence Specificity of Actinomycin D and Netropsin Binding to pBR322 DNA Analyzed by Protection from DNase I. *Proc. Natl. Acad. Sci. U.S.A.* 20, 3260–3264.
(17) Lee, M., Coulter, D. M., Pon, R. T., Krowicki, K., and Lown, J. W. (1988) Sequence Specific Molecular Recognition and Binding of a Monocationic Bis-Imidazole Lexitropsin to the Decadeoxyribonucleotide d-[(GATCCGTATG):(CATACGGATC)]: Structural and Dynamic Aspects of Intermolecular Exchange Studied by ^{1}H-NMR. *J. Biomol. Struct. Dyn.* 5, 1059–1087.
(18) Lee, M. S., Chang, D. K., Hartley, J. A., Pon, R. T., Krowicki, K., and Lown, J. S. (1988) Structural and Dynamic Aspects of Binding of a Prototype Lexitropsin to the Decadeoxyribonucleotide d(CGCAATTGCG)$_2$ Deduced from High-Resolution ^{1}H-NMR Studies. *Biochemistry* 27, 445–455.
(19) Lee, M., Shea, R. G., Hartley, J. A., Kissinger, K., Vesnaver, G., Breslauer, K. J., Dabrowiak, J. C., and (1989) Molecular Recognition Between Oligopeptides Acids: Sequence Specific Binding of the Naturall Antibiotic (4S)-(+) Anthelvencin A and its (4R)-(−) to Deoxyribonucleic Acids Deduced from ^{1}H-NMR: I Studies and Thermodynamic Data . *J. Am. Chen* 345–354.
(20) Lee, M., Shea, R. G., Hartley, J. A., Kissinger, K., V Breslauer, K. J., Pon, R. T., Dabrowiak, J. C., and I (1989) Molecular Recognition Between Oligopeptides . Acids: Sequence Specific Binding of (4S)-(+)- and (hydrokikumycin B to DNA Deduced from ^{1}H-NMR, F Studies and Thermodynamic Data. *J. Mol. Recogn.*
(21) Lee, M., Krowicki, K., Hartley, J. A., Pon, R. T., a W. (1988) Molecular Recognition Between Oligope Nucleic Acids: Influence of van der Waals Contact mining the 3′-Terminus of DNA Sequences Read by M Lexitropsins. *J. Am. Chem. Soc.* 110, 3641–3649.
(22) Rao, E. K., Yadigiri, B., and Lown, J. W. (1990) Recognition Between Oligopeptides and Nucleic Ac Sequence Specificity and Binding Properties of Thi tropsins Incorporating the Concepts of Site Accep Avoidance. *Anti-Cancer Drug Des.* (in press).
(23) Rao, K. E., Bathini, Y., & Lown, J. W. (1990) Synthes Thiazole-Containing DNA Minor Groove Binding Oli Related to the Antibiotic Distamycin. *J. Org. Chem.* 5
(24) Hélène, C. (1987) Lecture on Antisense Probes. I *tional Symposium on DNA as a Target for Chemot Action*, Oct 11–15, 1987, Seillac, France.
(25) Goodsell, D., and Dickerson, R. E. (1986) Isohelical . DNA Groove-Binding Drugs. *J. Med. Chem.* 29, 727
(26) Dervan, P. B. (1986) Design of Sequence Specific DN Molecules. *Science* 232, 464–471.
(27) Khorlin, A. A., Kroglov, A. S., Grokhovsky, S. L., Zh Zasedatelev, A. S., Gursky, G. V., and Gottikh, B. P. (19 Type of AT-Specific Ligand Constructed of Two Netr Molecules. *FEBS Lett.* 118, 311–314.
(28) Lown, J. W., Krowicki, K., Bahrarini, J., Newman, De Clercq, E. (1989) Novel Linked Antiviral and . Agents Related to Netropsin and Distamycin: Syn Biological Evaluation. *J. Med. Chem.* 32, 2368–2375.
(29) Dabrowiak, J. C., Skorobogaty, A., Rich, N., Varg, C. Vournakis, J. N. (1986) Computer Assisted Microden Analysis of Footprinting Autoradiographic Data. *Nuc Res.* 14, 489–499.
(30) Morgan, A. R., Lee, J. S., Pulleyblank, D. E., Murray, Evans, D. H. (1979) Ethidium Fluorescence Assays. Physicochemical Studies. *Nucleic Acids Res.* 7, 547–
(31) Le Pecq, J. B., and Paoletti, C. (1967) A Fluorescent Between Ethidium Bromide and Nucleic Acids. *J. Mol* 87–106.
(32) Youngquist, R. S., and Dervan, P. B. (1985) Sequenc Recognition of B-DNA by Oligo(N-methylpyrrolecarb *Proc. Natl. Acad. Sci. U.S.A.* 82, 2565–2569.
(33) Ward, B., Rehfuss, R., Goodisman, J., and Dabrow (1988) Determination of Netropsin–DNA Binding Const Footprinting Data. *Biochemistry* 27, 1198–1205.
(34) Ward, B., Rehfuss, R., and Dabrowiak, J. C. (1987) Qu Footprinting Analysis of the Netropsin–DNA Interac *Biomol. Struct. Dyn.* 4, 685–695.
(35) Pjura, P. E., Grzeskowiak, K., and Dickerson, R. Binding of Hoechst 33258 to the Minor Groove of B- *Mol. Biol.* 197, 257–271.
(36) Teng, M.-K., Usman, N., Frederick, C. A., and Wang (1988) The Molecular Structure of the Complex of Hoec and the DNA Dodecamer d(CGCGAATTCGCGC)$_2$. *Acids Res.* 16, 2671–2690.
(37) Griffin, J. H., and Dervan, P. B. (1987) Metalloregulati Sequence Specific Binding of Synthetic Molecules to I *Am. Chem. Soc.* 109, 6840–6842 and references therein

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.